PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-335144
(43)Date of publication of application : 17.12.1996

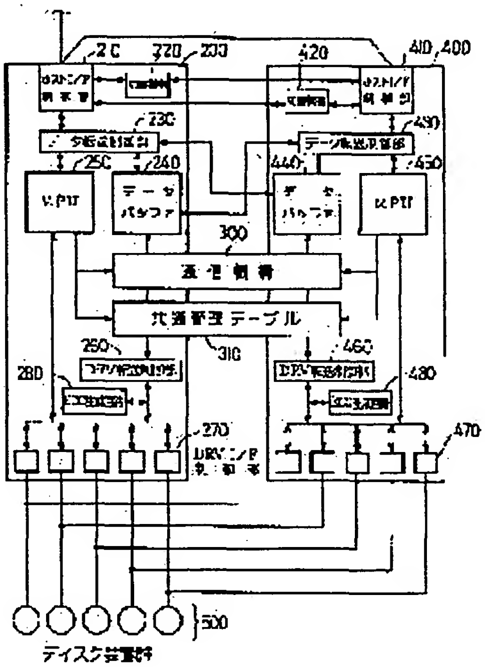
(51)Int.Cl. G06F 3/06
G06F 11/20
G06F 12/16
G06F 13/14

(21)Application number : 07-139781
(22)Date of filing : 07.06.1995

(71)Applicant : HITACHI LTD
(72)Inventor : MATSUMOTO YOSHIKO
MURAOKA KENJI

(54) EXTERNAL STORAGE DEVICE

(57)Abstract:
PURPOSE: To improve reliability and performance and to provide non-stop maintenance by distributing a load to the plural storage controllers of redundant configuration.
CONSTITUTION: Plural disk drive controllers 200 and 400 of redundant configuration for controlling a disk device are connected to a host device by the same SCSIID, the monitor of mutual operating states and the setting of load distribution information are performed by interposing a communication mechanism 300 and a common managing table 310 and in a normal state, high performance is provided by distributing the load by simultaneously operating the plural disk drive controllers 200 and 400 but in case of fault or maintenance, non-stop operation and non-stop maintenance are provided by executing a switching operation at the degeneracy and recovery caused by disconnection on the side of the fault while using switching mechanism 220 and 420.



LEGAL STATUS

[Date of request for examination] 06.06.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

1. This document has been translated by computer.So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] External storage containing two or more memory control units which intervene between the storage with which the data delivered and received between the high order equipment characterized by providing the following are stored, and the aforementioned storage and the aforementioned high order equipment, and control transfer of the aforementioned data between the aforementioned high order equipment and the aforementioned storage. An interface means to connect the memory control unit concerned to the aforementioned high order equipment so that two or more aforementioned memory control units may look equivalent [in view of the aforementioned high order equipment]. A surveillance means to be prepared in each aforementioned memory control unit, and to supervise the existence of the obstacle in other aforementioned memory control units, or change instructions. The change means which changes whether which aforementioned memory control unit controls transfer of the aforementioned data between the aforementioned high order equipment by being prepared in each aforementioned memory control unit. A communication-of-information means to transmit the mutual information of the aforementioned memory control unit, and a load-distribution means to make the load which originates in input/output request from the aforementioned high order equipment share among two or more aforementioned memory control units.
[Claim 2] External storage according to claim 1 characterized by providing the following. The data buffer which stores temporarily the aforementioned data which are prepared in each of two or more aforementioned memory control units, and are delivered and received between the aforementioned high order equipment. While writing write request data in alternative or multiplex to each of two or more aforementioned data buffers at the time of the write request from the aforementioned high order equipment Are at the write-in completion time to the aforementioned data buffer of the aforementioned write request data, and write in to the aforementioned high order equipment and completion is reported. The light after processing in which the aforementioned write request data are made to reflect from the aforementioned data buffer to the aforementioned storage asynchronously [the input/output request from the aforementioned high order equipment]. And they are the data transfer control means which can be performed alternatively about the write-through processing which it is at the write-in completion time to the aforementioned storage of the aforementioned write request data, and writes in to the aforementioned high order equipment and reports completion.
[Claim 3] External storage according to claim 1 or 2 characterized by providing the following. The 1st management information for being made accessible in common from two or more aforementioned memory control units, and discriminating whether each aforementioned memory control unit is healthy. The 2nd management information which specifies any shall be performed between the aforementioned light after processing and the aforementioned write-through processing. A management information storage means by which at least one of the 3rd management information which specifies any of two or more aforementioned memory control units receive the input/output request from the aforementioned high order equipment, and the 4th management information which specifies the assignment of the aforementioned load in each of two or more aforementioned memory control units is stored. The control logic carry out

operation of performing degeneracy operation which continues transfer of the aforementioned data between the aforementioned high order equipment by the aforementioned remaining memory control unit while separating the aforementioned memory control unit which the aforementioned obstacle occurred ignited by generating of an obstacle, or the change instructions from the outside, or was ordered from the outside, and operation return the separated aforementioned memory control unit to a redundant configuration.
[Claim 4] It is the external storage characterized by having the control logic which performs a halt and resumption of alternative write-in operation of the aforementioned write request data to each of the aforementioned data buffer by which the aforementioned data transfer control means were prepared in each of each aforementioned memory control unit in external storage according to claim 2.
[Claim 5] External storage characterized by performing maintenance of the micro program which controls the maintenance or the aforementioned memory control unit of a data buffer corresponding to the stopped aforementioned memory control unit while stopping at least one in two or more aforementioned memory control units alternatively and performing degeneracy operation in external storage according to claim 4.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] Especially this invention is applied to the external-memory subsystem of the redundant configuration which equipped multiplex with the input/output control unit which controls the input/output request of the information from high order equipment etc. about external storage, and relates to effective technology.
[0002]
[Description of the Prior Art] In the external storage which constitutes a computer system, when the memory control unit which intervenes between storage and high order equipment equipped with the storage, and controls transfer of the information between both is not a redundant configuration, if an obstacle occurs in a memory control unit, a subsystem will be obliged to a halt and a rehabilitation work will be performed in the meantime. And an end of this rehabilitation work resumes the business which the memory control unit was rebooted, or the subsystem was rebooted, and had been interrupted till then.
[0003] Moreover, recently, the employment gestalt of operation is increasing 24 hours in various information processing business which uses a computer system, and continuous running is demanded also of the external-memory subsystem. For this reason, for example, the technology whose continuation operation of a system one memory control unit tends to enable by the memory control unit of on stream and an other system stopping, and taking redundant composition about a memory control unit [say / taking a standby state], and changing to the memory control unit of the standby state of an other system at the time of the obstacle of a memory control unit is known as indicated by JP,3-206529,A.
[0004]
[Problem(s) to be Solved by the Invention] However, in the above-mentioned conventional technology, although continuous running at the time of an obstacle was possible, in spite of having had two sets of memory control units, actually working is only any one set and it did not change at all with the time of one set efficiently. That is, it was redundant, and also the memory control unit of a system could not but be an object for hot standbies to the last, and could not but be a mere alternative of the memory control unit of an obstacle.
[0005] Moreover, in recent years, the demand to a system was also various, there were various topologies by which an access demand is published from high order equipment to two or more paths to the same storage or separate storage, and it was difficult to build a system in the mere redundant configuration of a memory control unit like before according to various users' request.

[0006] Moreover, in the former, it has the composition that the memory control unit and the data buffer were carried in one board in the cheap system. When performing maintenance control, such as extension of the data buffer in a memory control unit, separation of only a data buffer Eye an impossible hatchet, A data buffer is extended in the state where the system was made to suspend. After the end of extension work, It was impossible to have carried out the maintenance control work of extension etc., having rebooted the memory control unit and the system, having completed the procedure of resuming the business interrupted till then, and processing the input/output request (I/O) from high order equipment.

[0007] The purpose of this invention is by making two or more memory control units of a redundant configuration distribute a load to offer the external storage which can raise reliability and a performance.
[0008] Other purposes of this invention are to offer the external storage which can realize improvement in the reliability by multiplexing of a memory control unit, and control action with a still more various memory control unit, without making it conscious of the redundant configuration of a memory control unit to a high order equipment side.
[0009] The purpose of further others of this invention is to offer the external storage which can carry out the maintenance control work of the hardware in two or more memory control units of a redundant configuration, software, etc. simple, without stopping operation.
[0010] The purpose of further others of this invention is to offer the external storage which can do the maintenance control work of composition of having carried the memory control unit and the data buffer on the single board during operation.
[0011]
[Means for Solving the Problem] An interface means by which the external storage of this invention connects a memory control unit to high order equipment so that two or more memory control units may look the same [in view of high order equipment], A surveillance means to supervise other memory control units among two or more memory control units, and the means of communications which can transmit the information between memory control units, It considers as composition including the change means which changes the memory control unit which has received the demand from high order equipment, the input/output request from the high order equipment which one memory control unit received, and the load-distribution means which carries out the load distribution of the processing which accompanies it in two or more memory control units.
[0012] Moreover, the data buffer which once stores the write-in data from high order equipment by preparing for each memory control unit and taking the same redundant configuration as a memory control unit, When the write-in data from high order equipment are stored in a data buffer, while reporting an end to high order equipment and writing the demand with high order equipment in storage from a data buffer asynchronously It considers as composition including a data transfer means to control whether it writes in two or more data buffers of all of a redundant configuration, or it writes in alternatively.
[0013] Moreover, the 1st management information for being made accessible in common from two or more memory control units, and discriminating whether each memory control unit is healthy, The 2nd management information which specifies any shall be performed between light after processing and write-through processing, It considers as composition including a management information storage means by which at least one of the 3rd management information which specifies any of two or more memory control units receive the input/output request from high order equipment, and the management information [of ** the 4th which specifies the assignment of the load in each of two or more memory control units] **s is stored.

[0014]
[Function] In the external storage of this invention, when it is a redundant configuration containing the 1st of a couple, and the 2nd memory control unit, daisy chain connection is made with for example, high order equipment and a SCSI interface, and these the 1st and 2nd memory control units are accessed by the same SCSIID, for example. For example, when the 1st memory control unit has received the input/output request from high order equipment fixed, other 2nd memory control unit distributes based on the 4th management information which specifies the assignment of the load in each of two or more memory control units, and the load accompanying processing of the input/output request concerned can aim at improvement in the throughput of the radial transfer by parallel operation of the improvement in the reliability by the redundant configuration, the 1st, and 2nd memory control units.
[0015] Moreover, he does not need to be [as opposed to / a change / that what is necessary is just to publish an I/O demand to SCSIID also with after / same / high order equipment] an

obstacle by changing the memory control unit which detects an obstacle by the 1st management information and the surveillance means when the 1st memory control unit has received the input/output request from high order equipment fixed, for example and an obstacle occurs in the 1st memory control unit, and receives a demand by the change means to the 2nd memory control unit] conscious at all. Then, the memory control unit acting as the obstacle is separated, and it goes into degeneracy operation. The 1st memory control unit is restored after the maintenance-service end of parts, exchange of a micro program, etc., and it is restored to the original redundant configuration.

[0016] Moreover, have a data buffer in each memory control unit, and the 1st memory control unit receives the write-in data from high order equipment. When light after processing is being performed, a surveillance means detects that the obstacle occurred in the 1st memory control unit. by the change means When the memory control unit which receives input/output request is changed from the 1st memory control unit to the 2nd memory control unit, Simultaneously, it changes from the multiplex data write-in processing to two or more data buffers to the processing which writes data in the data buffer with which the memory control unit under operation was equipped alternatively.

[0017] At this time, it chooses whether light after processing is performed or write-through processing is performed. This selection is possible when a user sets up the 2nd management information of a management information storage means. That is, when the demand to a user's data reliability is high, and setting it as write-through mode and requiring a performance rather than reliability, it is set as light after mode.

[0018] It changes to the operation written in a data buffer by changing to alternative writing or multiplex writing multiplex after restoration of the 2nd memory control unit, and can restore to a redundant configuration.

[0019] After the 1st memory control unit receives the input/output request from high order equipment, and changing from the processing written in multiplex when double writing is performed to the data buffer of the 1st memory control unit, and the data buffer of the 2nd memory control unit and light after processing is being performed about the write request from high order equipment to the processing written in alternatively, separating the 2nd memory control unit, degenerating it and maintaining extension of a data buffer, exchange of a micro program, etc., the original redundant configuration is restored. Then, the 2nd memory control unit for means of communications of **** which can transmit the information between memory control units notifies completion of maintenance services, such as extension of a data buffer, to the 1st memory control unit, and changes receipt of the demand from high order equipment to self-equipment after a notice using a change means.

[0020] On the other hand, the 1st memory control unit which received the notice degenerates self-equipment, maintains extension of a data buffer etc. and makes it restore. The 1st memory control unit notifies the completion of maintenance of extension of a data buffer etc. to the 2nd memory control unit after restoration using the means of communications which can transmit informational. Ignited by this, it changes from the alternative data writing to a single data buffer to the multiplex write-in processing to two or more data buffers. Thereby, it becomes possible, continuing the radial transfer [business / maintenance control /, such as extension of the data buffer of the 1st memory control unit of a couple, and the 2nd memory control unit, and exchange of a micro program,] between high order equipment.

[0021] Moreover, according to this invention, the 1st memory control unit and the 2nd memory control unit can judge any receive the demand from high order equipment by referring to the 3rd management information which specifies any of two or more aforementioned memory control units set as the management information storage means receive the input/output request from the aforementioned high order equipment. It is also possible for this to receive and process input/output request in the memory control unit of either the 1st memory control unit or the 2nd memory control unit not only receiving the demand from high order equipment but both. Moreover, it becomes possible by setting up the 3rd management information optionally by the user to specify from a user the memory control unit which receives the demand from high order equipment to be arbitration.

[0022]

[Example] Hereafter, the example of this invention is explained in detail, referring to a drawing.

[0023] Drawing 1 is the conceptual diagram showing an example of the computing system containing the external storage which is one example of this invention. The computing system of this example contains the high order equipment 100 which is a central processing unit, the disk drive control unit 200 and the disk drive control unit 400, and the disk unit 500. The disk drive control unit 200 and the disk drive control unit 400 were connected with high order equipment 100 with the daisy chain of a SCSI interface, the same SCSIID was set up and the disk drive control unit 200,400 has taken the redundant configuration. And in the case of this example, the disk drive control unit 200 receives the demand from high order equipment 100, performs processing which accompanies a demand with the disk drive control unit 200 and the redundant disk drive control unit 400, and controls a disk unit 500. [0024] Drawing 2 is the block diagram showing an example of the internal configuration of the disk drive control units 200 and 400. In addition, since the internal configuration of the disk drive control unit 200 and the disk drive control unit 400 is the same, the disk drive control unit 200 is explained to an example, 2 figures is made the same under the sign of the part which corresponds about the disk drive control unit 400 side, and explanation is omitted.

[0025] a microprocessor unit 250 (Following MPU is called) is performed decoding a RAM (RAM -- not shown) serially, and is controlling the whole disk drive control unit 200

[0026] The host I/F control section 210 is performing protocol control with high order equipment 100. The DRVI/F control section 270 is performing protocol control with each drive. A data buffer 240 is used at the time of the data transfer of the host I/F control section 210 and the DRVI/F control section 270. Volatilization memory is sufficient as this memory, and non-volatilized memory is sufficient as it. this example describes for an example the case where a data buffer 240 is built by volatilization memory.

[0027] The change mechanism 220 is for changing the host I/F control section which receives I/O from high order equipment 100 to the host I/F control section 210 and the host I/F control section 410 of each disk drive control unit 200 and the disk drive control unit 400. The host I/F control section 210 shall have received in this example. The data transfer control section 230 is controlling the data transfer of high order equipment 100 and a data buffer 240. It has the function of whether this data transfer control section 230 carries out double writing of the light data from high order equipment 100 to the 2nd page, a data buffer 240 and a data buffer 440, or to 1-fold carry out [of only a data buffer 240 / both] writing. Moreover, it is possible to change 1-fold writing or double writing with the directions from MPU250.

[0028] The DRV transfer control section 260 controls the data transfer between a data buffer 240 and a disk unit 500.

[0029] The transmitter style 300 is a mechanism for transmitting the information between MPU250 and MPU450. This transmitter style 300 is enabling the bidirectional transfer between MPU250 and MPU450.

[0030] The common managed table 310 is a managed table in which both refernce / renewal of MPU250 and MPU450 are possible.

[0031] this example takes and explains to an example the drive storing method by array composition which is distributed to two or more disk units 500, and stores the logical data from high order equipment 100.

[0032] The ECC generation circuit 280 has the function which generates redundant data to the data sent from high order equipment 100, and can use this function also for the reconstitution of data. 1 logical data unit sent from the high order is sufficient as the unit which adds redundant data, and it is good even to two or more logical data units. this example adds redundant data to four logical data, and describes them in RAID5 method which does not fix the drive which stores this redundant data.

[0033] Next, with reference to drawing 3 , an example of the composition of the common managed table 310 is explained. It is used for surveillance intelligence 320 confirming whether each disk drive control units 200/400 are operating normally. Surveillance intelligence A321 sets up information at a fixed interval, when MPU250 of the disk drive control unit 200 is normally judged that operation is possible. Moreover, when MPU250 judges normally that operation is impossible, the information which shows abnormalities is set up. In addition, MPU450 of the disk drive control unit 400 as well as MPU250 sets information as surveillance intelligence B322.

[0034] The data transfer mode information 330 directs the end report opportunity to the light data write request from high order equipment 100 at the time of the degenerate state of a system. That is, it is this information at the write-in completion time to a data buffer 240 or a data buffer 440, and is the information for judging whether an end is reported to high order equipment 100, or an end report is carried out when it writes even in a disk unit 500 from a data buffer 240 (write-through mode being called below) (light after mode being called below).

[0035] The directions information on a disk drive control unit that the host I/O receipt information 340 receives I/O between two disk drive control units 200/400 is shown. this example explains as that by which the disk drive control unit 200 is set as the host I/O receiving side.

[0036] The load-distribution information 350 is information for carrying out the load distribution of the processing accompanying I/O from high order equipment between [of two] disk drive control unit 200 / 400. The method of a load distribution may divide into each disk drive control unit the disk unit made applicable to access, and may share it with the processing which stores light data in a disk unit 500 from a data buffer with asynchronous processing of the I/O demand from high order equipment 100 and I/O demand from high order equipment 100. Or the method of performing processing is sufficient as the way which writes in all the matters that must be processed into load-distribution information, considers as competition logic between two MPU, and has an opening as MPU.

[0037] By this example, processing of the I/O demand from high order equipment 100 and the I/O demand from high order equipment 100 explain the method shared with the processing which stores light data in a disk unit 500 from data buffers 240/440 asynchronously. Therefore, in this example, the information on the light data stored in data buffers 240/440 shall be contained in the load-distribution information 350.

[0038] Next, the write-in processing and reading processing of data to a disk unit 500 are explained from the high order equipment 100 in the computing system in this example.

[0039] Usually, at the time of the write request from high order equipment 100, by the host I/F control section 210, the disk drive control unit 200 receives write-in logical data, stores it in a data buffer 240 and a data buffer 440 doubly by the data transfer control section 230, sets storing information as the load-distribution information 350 on the common managed table 310, and reports an end to high order equipment 100 at this time. Serially, if MPU450 has storing information with reference to the load-distribution information 350 The light data concerned and the data of the same address already stored in the drive (the old data are called below), The parity data corresponding to the light data concerned are read from a disk unit 500 by the DRVI/F control section 470 and the DRV transfer control section 460. Light data, the old data, and parity data generate the parity data (new parity data are called below) corresponding to light data in the ECC generation circuit 480. Light data are stored in a disk unit 500 by writing the new parity data and light data which were generated in a disk unit 500 by the DRVI/F control section 470 and the DRV transfer control section 460. This processing is asynchronously performed with the I/O demand from high order equipment 100. Moreover, the read-out processing of the old data / old parity data performed since light data are stored and new parity generation processing, and new parity data storage processing are called light penalty in RAID5.

[0040] Thus, the storing demand of the light data from high order equipment 100 is processing that a load is very high, when operating two or more disk units 500 as disk array equipment. Efficiency leads to the improvement in a performance as a system well rather than carrying out a role assignment and performing this processing with two disk drive control units 200,400 performs by one set only of a disk drive control unit. A cheap processor is carried especially as latest commercial-scene trend, and it has become a very important element with high performance and high-reliability to reduce system-wide cost. Therefore, in light penalty processing, although that much accesses to a drive occur also leads to performance degradation, since the transit time of the micro program of the processor which controls it before it is long, a

processor neck has many bird clappers as a system. At this time, the performance near the double precision can be taken out with processing by two sets of the disk drive control units 200 and 400 like this example.

[0041] Next, the reading demand from high order equipment 100 and MPU250 start reading of data from a physical drive (disk unit 500) by the DRVI/F control section 270 and the DRV transfer control section 260, and transmit it to high order equipment 100. Moreover, while the lead demand address from high order equipment 100 is continuing at this time, it may judge that the disk drive control unit 400 is sequential lead processing, and processing which reads asynchronously [I/O from high order equipment 100] an certain amount of data which follows the lead demand address from high order equipment 100 to data buffers 240 and 440 may be performed. When there is next an I/O demand from high order equipment by carrying out like this, the target data are already stored in data buffers 240/440, and data can be transmitted without producing access to the disk unit 500 which time requires, and it leads to the improvement in a performance as the whole.

[0042] As mentioned above, though it is a redundant configuration, it leads not only to reliability but to improvement in a performance by performing a part of processing rather than making a redundant portion (this example disk drive control unit 400) only stand by as an object for the change at the time of obstacle generating.

[0043] Next, in this example, while two sets of the disk drive control units 200/400 perform processing, operation which performs automatic switching and restoration at the time of an obstacle is explained. First, the surveillance procedure which detects an obstacle automatically is explained.

[0044] MPU 250 and 450 sets the information (normal information is called hereafter) which shows that MPU450 of MPU250 is normal to surveillance intelligence 321 whenever fixed time passes controlling the disk drive control units 200 and 400 as surveillance intelligence 322. However, in order to show having set up for every fixed time, the information which changes serially is set to this information. For example, it is the information which is added one [at a time]. Moreover, when accessing a data buffer is that each MPU250,450 judged normally that operation was impossible with the disk drive control unit 200,400 concerned impossible from MPU, for example, the information (this is called obstacle information below) which shows that it is an obstacle is set as surveillance intelligence. Hereafter, the flow chart of drawing 4 explains an example of the above-mentioned surveillance procedure.

[0045] Here, MPU450 of the disk drive control unit 400 takes and explains to an example operation which supervises the disk drive control unit 200 of an other system.

[0046] MPU250 judges first whether fixed time passed at Step 600. If fixed time has not passed, it progresses to Step 608 and it is judged that the disk drive control unit 200 is normal.

[0047] If fixed time has passed, it will progress to Step 601 and the normal information which shows that MPU450 is normal will be set up. And it progresses to Step 602 and the surveillance intelligence 322 of the disk drive control unit 200 is referred to. If it judges whether this information is normal and judges that it is normal at Step 603, it will progress to Step 604. If it judges that it is an obstacle, it will progress to Step 607 and it will be judged that the disk drive control unit 200 is an obstacle.

[0048] At the time of normal information, it progresses to Step 605 and judges at Step 605 whether this normal information had change from before. That is, MPU250 may have fallen impossible [a setup of surveillance intelligence] according to the obstacle of a micro program etc. Such an obstacle is judged with the check of this step 605. If there is change, it will progress to Step 608 and it will be judged that it is normal. When there is no change, it progresses to Step 606 and judges whether the time of a margin longer than fixed time has passed. Consequently, if it has passed, it progresses to Step 607 and is judged as an obstacle, and if it has not passed, it will progress to Step 608 and it will be judged that it is normal. According to the above surveillance procedure, both of obstacles of a micro program can also detect the obstacle of hardware simultaneously.

[0049] Next, an example of the processing from which the disk drive control unit 400 recognizes the obstacle of the disk drive control unit 200 of an other system, and changes with reference to the flow chart of drawing 5 is explained.

[0050] Refer to the load-distribution information 350 for MPU450 serially at Step 700 first. Consequently, if the light data from high order equipment 100 do not exist in a data buffer 240 and 440 at Step 701, it progresses to Step 704. If it exists, in order to progress to Step 702 and to generate the parity corresponding to the light data of a data buffer 440, the old data and the old parity data corresponding to the light data concerned are read from a disk unit 500, and new parity data are generated in the ECC generation circuit 480. Then, it progresses to Step 703 and light data and new parity data are stored in a disk unit 500 by the DRV transfer control section 460 and the DRVI/F control section 470. Next, at Step 704, the obstacle of the disk drive control unit 200 is checked in the surveillance procedure after Step 600 of drawing 4 . Consequently, if normal, it will progress to Step 700 and processing will be continued. If it judges that a change is required, it will progress to Step 710 and reception of I/O from high order equipment 100 will be changed from the disk drive control unit 200 to the disk drive control unit 400 using change procedure. And the disk drive control unit 400 substitutes Step 720 for the I/O processing from high order equipment 100 which the disk drive control unit 200 was performing, and it is performed.

[0051] Next, it changes with the flow chart of drawing 6 , and an example of procedure is explained.

[0052] At Step 711, it directs to write the data concerned in a data buffer 440 one-fold to the data transfer control section 430 at the time of the light data receipt from high order equipment 100 first. That is, since a data buffer does not exist in the disk drive control unit 400 until it exchanges the part which the disk drive control unit 200 was degenerated, separated, and carried out obstacle generating and restores, since the obstacle occurred in the disk drive control unit 200, double writing like [at the time of a normal redundant configuration] is not made.

[0053] And it directs to change the I/O demand from high order equipment 100 from the host I/F control section 210 to the host I/F control section 410 by the change mechanism 420 by Step 712. Although it stops receiving the demand from high order equipment 100, as for the host I/F control section 210, the host I/F control section 410 will come to receive the demand from high order equipment 100 and a disk drive control unit will change substantially by this, at this example, there is no need of knowing the disk drive control unit by the side of receipt having changed that SCSIID should just publish eye the same hatchet and high order equipment 100 to SCSIID same before changing I/O.

[0054] Next, after changing using the flow chart of drawing 7 , an example of a procedure which performs I/O with the disk drive control unit 400 is explained.

[0055] If I/O processing is received from high order equipment 100 at Step 721, it will progress to Step 722 and a lead demand or a light demand will be judged. At the time of a lead demand, it progresses to Step 729 and object data are read into a data buffer 440 from the disk unit 500 corresponding to the lead demand concerned. It progresses to Step 730, data are transmitted to high order equipment 100 from a data buffer 440, and Step 728 reports an end to high order equipment 100.

[0056] At the time of a light demand, it progresses to Step 723 and light data are stored in a data buffer 440. Furthermore, it progresses to Step 724 and judges whether it is write-through mode at Step 725 with reference to the data transfer mode information 330. Consequently, it progresses to Step 728 at the time of the mode in which an end is reported to high order equipment 100 when stored at the time 440 of light after mode, i.e., a data buffer, it reports an end, and stores it in a disk unit 500 from a data buffer 440 asynchronously after that. It progresses to Step 726 at the time of write-through mode, it creates the parity data to light data, stores light data and new parity data in a disk unit 500 at Step 727, and reports an end at Step 728. Furthermore, after this, Step 703 is performed from Step 700 in the flow chart of drawing 5 , and processing before a change is also performed.

[0057] Thus, according to this example, mutual change operation and continuation of processing are automatically possible without the directions from high order equipment 100 at the disk drive control units 200/400, without carrying out no consciousness to high order equipment 100.

[0058] Next, the disk drive control unit 200 is restored and an example of the method when returning to the original redundant configuration is explained.

[0059] First, the flow chart shown in drawing 8 explains an example of restoration operation by the side of the disk drive control unit 200. It notifies that restoration was completed to the disk drive control unit 400 by Step 810 at transmitter guard 300. Then, the disk drive control unit 200 turns into a redundant disk drive control unit, and before and a position interchange. At Step 811, asynchronous DESUTEJI processing (Steps 700-705 of drawing 5) which the disk drive control unit 400 was performing is performed before.

[0060] Furthermore, an example of operation of the near disk drive control unit 400 which received the notice with the flow chart shown in drawing 9 is explained.

[0061] If the completion of restoration of the disk drive control unit 200 is recognized at transmitter guard 300 by Step 820, it will point to double writing to data buffers 240 and 440 to the data transfer control section 430 at Step 821, and only I/O processing from high order equipment 100 will be performed at Step 821. Thus, receiving the I/O demand from high order equipment 100, restoring in the original redundant composition is possible, and improvement in a performance can also be aimed at by carrying out the load distribution of the processing with two more disk drive control units 200/400.

[0062] Next, it explains, referring to the flow chart of drawing 10 about an example of the extension method of the data buffer under operation of the disk drive control units 200/400. In addition, let the disk drive control unit which has received I/O from high order equipment 100 be the disk drive control unit 200.

[0063] When there is an extension demand of a data buffer, the disk drive control unit concerned judges whether it is an I/O receiving side at Step 911. The content of processing of the disk drive control unit 200 is explained first. Since the disk drive control unit 200 is an I/O receiving side, it progresses to Step 912, the disk drive control unit 400 degenerates first, and it recognizes separating. Then, it directs to make light data into 1-fold writing to a data buffer 240 to the data transfer control section 230. Then, I/O processing from high order equipment 100 is performed at Step 913, and Step 700 of drawing 5 - Step 703 are performed at Step 914. That is, it substitutes for a part to have performed with the disk drive control unit 400. It waits for the completion of restoration of the disk drive control unit 400, repeating Step 913 and Step 914.

[0064] Next, as for the disk drive control unit concerned, the disk drive control unit 400 also judges whether it is an I/O receiving side at Step 911. Since it is not an I/O receiving side as a result, it progresses to Step 915, and the disk drive control unit 400 concerned detaches, and extends a data buffer 440 at Step 916. It notifies having restored at Step 917 to the disk drive control unit 200 through the transmitter style 300 after the completion of extension.

[0065] Since the disk drive control unit 200 needs to extend shortly, the disk drive control unit 400 changes the host I/F control section which carries out I/O reception using the change mechanism 420 at Step 919 to self-** in order to substitute for receipt of I/O. Then, I/O processing from high order equipment 100 is performed at Step 920, Step 700 of drawing 5 - Step 703 are performed at Step 921, and it waits for the completion of restoration of the disk drive control unit 200.

[0066] The disk drive control unit 200 which has recognized restoration through the transmitter style 300 at Step 918 is separated at Step 922, and extends a data buffer 240 at Step 923. The transmitter style 300 notifies restoration to the disk drive control unit 400 at Step 924 after the completion of extension. After a notice, since the disk drive control unit 200 concerned is not a host I/O receiving side, it turns to the side which performs Step 700 of drawing 5 - Step 705 at Step 925.

[0067] The disk drive control unit 400 will direct to write light data to data buffers 240/440 doubly to the data transfer control section 230 at Step 927, if restoration of an other system is recognized at transmitter guard 300 by Step 926. I/O processing from high order equipment 100 is performed at Step 928.

[0068] Thus, though I/O from high order equipment 100 is performed, extension of the data buffers 240/440 of each ** is attained. That is, according to this example, by the former, extension of extension of a data buffer is attained in online to having been unrealizable unless it was after suspending a system. Especially when the disk drive control unit was built on one board realized in the low cost, the exchange for every board was impossible for extension under eye a required hatchet and operation. In this example, extension of a data buffer is possible, setting to the disk drive control units 200/400 of a redundant configuration, and degenerating / restoring one set at a time.

[0069] Moreover, according to this example, by transposing processing of Step 916 of drawing 10 , and Step 923 to micro program exchange work, exchange of the micro program under operation is possible, and the demand of 24-hour operation is effective in especially the maintenance control work in a remarkable computer system in recent years.

[0070] Moreover, a user can direct whether for the light demand from high order equipment 100 to be written in by the data buffer, and to report an end, or to write even in a disk unit 500 and report an end during the degeneracy at the time of a piece system obstacle. That is, you may perform automatically rewriting of this data transfer mode information 330 by a user's program. That is, if an end is reported when a data buffer becomes 1st page composition, and stored in data FABBA, although it excels in responsibility, a data guarantee becomes impossible when an obstacle occurs in a disk drive control unit at this time. Since light penalty processing occurs in on the other hand storing even in a disk unit 500, although responsibility will deteriorate considerably, a positive response can be reported to high order equipment 100, and it is reliable. In the case of the external storage of this example, according to the demand level of the reliability to the file which a user treats, it can choose optionally whether priority is given to reliability, or priority is given to a speed of response with directions of a user, and it becomes possible to build a flexible file system.

[0071] Furthermore by this invention, two or more disk drive control units can also offer simultaneously the system which can be accessed not only from redundant composition but from two or more high order equipment or two or more buses. This example of a system configuration is shown in drawing 11 and drawing 12 .

[0072] Although drawing 11 is the same composition as drawing 1 of an example explained until now, when I/F with high order equipment 100 is SCSI, with the composition of drawing 1 , the points connected by SCSIID from which a memory control unit 0 (400A) and a memory control unit 1 (200A) differ differ by the composition of drawing 11 to the memory control unit 0 and the memory control unit 1 having been connected by the same SCSIID. In the composition of this drawing 11 , both receive and process an I/O demand from high order equipment 100. Moreover, drawing 12 is the block diagram showing an example of the system configuration to which two or more memory control units 0 (400B) and memory control units 1 (200B) were connected by the multi-pass to the same high order equipment 100. The composition of this drawing 12 of a memory control unit 0 (400B) and a memory control unit 1 (200B) is all also an execute permission about the I/O demand from high order equipment 100. Specification of any perform an I/O demand is realized by rewriting the host I/O receipt information 340 of the common managed table 310. That is, each memory control unit determines first whether the memory control unit concerned receives I/O from high order equipment with reference to the host I/O receipt information 340. Thus, in this invention, it can respond to various users' connection method, and a flexible system can be built.

[0073] As explained above, while two or more disk drive control units 200/400 of a redundant configuration carry out a load distribution, according to this example, the file system which can realize simultaneously not only the improvement in reliability but improvement in a performance can be offered by performing the demand from high order equipment 100. Moreover, performing the I/O demand from high order equipment 100, while all the disk drive control units 200/400 carry out a load distribution, but, it changes automatically, operation is continued, without looking for directions in any way from high order equipment 100 at the time of obstacle generating, and it becomes possible to restore further. It becomes exchangeable [extension of a data buffer, or a micro program] by this, performing the I/O demand from high order equipment 100, and non-stopped maintenance can be realized. Moreover, not only a redundant configuration but all disk drive control units are possible also for making it the composition which receives the demand from high order equipment 100 simultaneously, and can respond to the various file systems

which a user demands flexibly.

[0074]

[Effect of the Invention] According to the external storage of this invention, the effect that reliability and a performance can be raised is acquired by making two or more memory control units of a redundant configuration distribute a load.

[0075] Moreover, the effect that improvement in the reliability by multiplexing of a memory control unit and control action with a still more various memory control unit are realizable is acquired, without making it conscious of the redundant configuration of a memory control unit to a high order equipment side.

[0076] Moreover, the effect that the maintenance control work of the hardware in two or more memory control units of a redundant configuration, software, etc. is executable simple is acquired, without stopping operation.

[0077] Moreover, the effect that the maintenance control work of composition of having carried the memory control unit and the data buffer on the single board can be done during operation is acquired.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the conceptual diagram showing an example of the computing system containing the external storage which is one example of this invention.
[Drawing 2] It is the block diagram showing an example of the internal configuration of the disk drive control unit which constitutes the external storage which is one example of this invention.
[Drawing 3] It is the conceptual diagram showing an example of the composition of the common managed table used in the external storage which is one example of this invention.
[Drawing 4] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 5] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 6] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 7] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 8] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 9] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 10] It is the flow chart which shows an example of an operation of the external storage which is one example of this invention.
[Drawing 11] It is the conceptual diagram showing the modification of a topology with the high order equipment in the external storage which is one example of this invention.
[Drawing 12] It is the conceptual diagram showing the modification of a topology with the high order equipment in the external storage which is one example of this invention.
[Description of Notations]
100 [-- Host I/F control section,] -- High order equipment, 200 -- A disk drive control unit, 210 220 [-- Data buffer,] -- A change mechanism, 230 -- A data transfer control section, 240 250 -- A microprocessor unit, 260 -- DRV transfer control section, 270 [-- Transmitter style,] -- An DRVI/F control section, 280 -- An ECC generation circuit, 300 310 [-- Surveillance intelligence,] -- A common managed table, 320 -- Surveillance intelligence, 321 322 [-- Host I/O receipt information,] -- Surveillance intelligence, 330 -- Data transfer mode information, 340 350 [-- Host I/F control section,] -- Load-distribution information, 400 -- A disk drive control unit, 410 420 [-- A data buffer, 450 / -- A microprocessor unit, 460 / -- A DRV transfer control section, 470 / -- An DRVI/F control section, 480 / -- An ECC generation circuit, 500 / -- Disk unit.] -- A change mechanism, 430 -- A data transfer control section, 440

[Translation done.]